



A JOURNAL OF THE SOCIETY FOR MUSIC THEORY

Prosodic Dissonance*

Eron Smith

NOTE: The examples for the (text-only) PDF version of this item are available online at:

<https://www.mtosmt.org/issues/mto.24.30.2/mto.24.30.2.smith.php>

KEYWORDS: prosody, accent, stress, text setting, enjambment, Kesha, Royal & the Serpent, Rina Sawayama

ABSTRACT: In popular and scholarly discourse on texted music and music-speech intersections, the prevailing assumption is that the linguistic features of the lyrics (e.g., syllabic stress) align with the way the lyrics are sung (e.g., phenomenal accent in or affecting the melody)—or, if not, that they should. However, not only *can* text and music conflict, but they do so in a variety of ways, to varying degrees, and with different effects on our listening experience. I define *prosodic dissonance* as any conflict between the prosodic linguistic features and musical rendition of text. This could include misalignment between syllabic and durational/registral stress, between spoken and sung phrases, or between spoken and sung intonation. Prosodic dissonance/consonance can also interact with rhyme, vowel shape, parallelism, and syncopation. To recognize prosodic dissonance, I (1) determine the prosody for the lyrics as spoken, (2) determine the prosody for the melody as sung, (3) identify mismatches as dissonances, (4) consider the effect of the surrounding melody/lyrics, (5) consider alternate pronunciations or hearings that might account for it, and (6) consider the perceptual and analytical implications. This article focuses on prosodic dissonance in popular music, with longer analyses of Kesha’s “Tonight” (2020), Royal & the Serpent’s “Overwhelmed” (2020), and Rina Sawayama’s “This Hell” (2022).

DOI: 10.30535/mto.30.2.6

Received April 2023

Volume 30, Number 2, June 2024
Copyright © 2024 Society for Music Theory

1. Introduction

Unconditionally Rejected

[1.1] Katy Perry released her 2013 single “Unconditionally” to mixed reviews.⁽¹⁾ Some critics focused on cultural appropriation in live performances (e.g., Feeney 2013), some connected the song’s sound to their Christian beliefs and the artist’s Christian music roots (e.g., Lynch 2016), and others noted its “self-help-y lyrics” (Brown 2013). In reviews and across the internet more broadly, however, one particular feature of the song has received consistent—and consistently *negative*—attention: the singer’s apparent delivery of the title as “uncondiTIONal.” Professional and amateur

critics across the internet noted Katy Perry's rendition of "unconditional," extracted in **Audio Example 1**, as weird, ridiculous, laughable, and wrong. For some listeners, this setting went so far as to "ruin" the song entirely.⁽²⁾ Despite songwriting credits from four experienced songwriters (Katy Perry, Dr. Luke, Max Martin, and Cirkut), the song's reception framed the chorus as containing a "mistake."

[1.2] While negative value judgments are common in internet responses to pop music releases, the singular focus on the title word reveals an attached set of music-theoretical takeaways. When reviewers and writers criticize "emphasis on the wrong syllable" in song rather than in speech (also referred to as "bad text setting"), the logical progression of this argument is that

1. the text has a stress pattern of its own,
2. the musical setting has a stress pattern of its own,
3. these two stress patterns are incongruous, and
4. this incongruity produces the aural impression of mispronunciation of the text.

[1.3] The consistent responses of amateur and professional critics alike point to the important interactions between music and the musical qualities of language itself. This article investigates this interaction, focusing on the *dissonance and consonance between musical and linguistic prosody*. Before combining the two, let us establish some of the prevailing understandings of prosody, accent, and text-setting as they relate to speech and music—matters explored at length in music theory, cognitive science, and linguistics.⁽³⁾

Defining and Perceiving Stress and Its Role in English

Stress has always seemed to resist all attempts at definition: the closer one investigates the nature of stress, the more difficult it becomes to define. One of the reasons for this is that a stressed (or prominent) syllable is distinguished not only by acoustic features, but it is also a perceptual phenomenon, i.e., in defining it, one must account not just for its production, but also for its perception by the interlocutor. (Frost 2011, 68)

[1.4] While music theorists tend to discuss rhythm in conjunction with meter and accent, linguists often frame speech rhythm in terms of stress.⁽⁴⁾ Spoken English differentiates between stressed and unstressed syllables, tending to use more pitch movement, more distinct vowel qualities, and longer durations for stressed syllables. Historically, English has been categorized as a "stress-timed" language, tending to evenly distribute stressed syllables in time regardless of the number of syllables between them (as opposed to "syllable-timed" languages such as Spanish or French, which more evenly distribute syllables regardless of stress).⁽⁵⁾

[1.5] Though the stress-timed/syllable-timed dichotomy is widespread, some scholars have disputed it, proposing alternative ways of describing languages' stress patterns. Bertinetto (1989) provides a summary of this debate and offers "compressibility" as an improved means of distinguishing syllable stress distribution. Arvaniti (2009) argues that research on speech rhythm should be based on the same principles for all languages, proposing a focus on grouping and prominence. Cross-linguistic metric differences are not robust, and metric differences between different speakers and dialects of the same language can be substantial (Arvaniti 2012, Bertinetto 2021). Nevertheless, stress does seem to be a widely perceptible phenomenon, even in species other than humans (Hoeschele and Fitch 2016), and despite dwindling evidence in favor of clear stress-timed/syllable-timed language categories, it does seem to vary between languages. Although individual speakers of a language may use a wide variety of stress metrics, stress can still inform whether we hear speech as foreign (Patel 2010), and our native language can have both contextual and perceptual effects on how we hear stress, as in the "stress deafness" effect observed in native French speakers (Degrave 2019, Frost 2011).

[1.6] In short, stress is a language-specific phenomenon made up of multiple acoustic factors, which enables the distinction of homonyms and the segmentation of words and phrases, facilitating our

understanding (Cutler 2008). Since stress is a particularly important aspect of understanding and speaking English, the acoustic factors mentioned above—pitch movement, vowel quality, and duration—would intuitively seem more influential in English-language music than they would be in languages without such an emphasis on stress.⁽⁶⁾ This intuition, however, is dependent on the assumption that musical stress (particularly melodic stress) and linguistic stress have mutual influence.

Between Speech and Song

[1.7] Some of the evidence for permeability between musical and linguistic features in texted music comes from the categories of song and speech themselves. Song and speech—much like syllable- and stress-timed languages—are better described as a continuum than a dichotomy. The distinction between the two, as well as the categorization of speech and song as such, is culturally specific.⁽⁷⁾ Even within the same culture, the same sounds may be perceived as sung, spoken, or some combination of both, based on pitch stability, duration, timbre, rhythmic regularity, repetition, and context. Diana Deutsch's famous "sometimes behave so strangely" illusion (2019) illustrates this phenomenon, as does the social media trend of harmonizing unaltered speech with jazz transcriptions (Liberman 2021).

[1.8] Furthermore, some cultures do not distinguish between song and speech categories (Trehub, Becker, and Morley 2015), and even in cultures with such a distinction, some forms of expression occupy a gray area between them. Some art forms situated between song and speech originate thousands (e.g., religious chant and poetry) or hundreds of years ago (e.g., recitative). Others are more recent developments that combine song and speech, like the mutual merging of singing and rap (Komaniecki 2019; 2020), and the "orationality" of speech-like singing in emo and punk music (Chiu and Howie 2022), which can serve as generic and formal markers as well as means of emotional expression.

[1.9] Even when speech and song are clearly distinguished, each still affects the perception of the other. Various effects of musical ability and musical context on prosody have been documented, such as links between the effects of amusia and prosody perception (Hausen et al. 2013), rhythmic cues establishing expectations about speech prosody (Cason, Astésano, and Schön 2015), effects of musical familiarity on spoken delivery of lyrics (Albert 2021), and links between familiarity of musical excerpts and prosody (Palmer, Jungers, and Jusczyk 2001). Furthermore, musical training improves the perception of lexical stress, even for speakers of "stress-deaf" languages (Degrave 2019, Kolinsky et al. 2009). The link also holds in reverse: prosody may inform how we hear meter and syncopation, affect ease of performance (Reed, Maxwell, and Temperley 2019), or affect our ability to understand the words (Gordon, Magne, and Large 2011).⁽⁸⁾

Speech-song Interactions in Compositional and Analytical Practice

[1.10] In addition to studying the perception of language, music theorists and linguists have investigated the effects of language on compositional practice. Language and dialect have repeatedly been shown to correlate with text setting practices and poetic meters in both texted and untexted music. Some authors describe this correlation in terms of meter-accent alignment probabilities (Temperley and Temperley 2013), while others posit that nationality affects composers' preference for certain poetic meters or rhythmic practices (Rothstein 2008, Vukovics and Shanahan 2020, Daniele and Patel 2013). Generally, even while examining to what degree linguistic stress aligns with meter, most of these studies assume that stressed syllables align with musical accent.⁽⁹⁾

[1.11] Though music and speech differ in communicative capacity and degree of periodicity, both "involve the systematic temporal, accentual, and phrasal patterning of sound"—i.e., rhythm—as well as syntactical structures that group into higher-level units (Patel 2010, 176). In both media, this patterning and grouping involves varying pitch and duration (Wennerstrom 2001, Lerdahl and Jackendoff 1983). Likely owing to the commonalities described above, some similar hierarchical analytical conceptions and visual representations of meter have been used for music, poetry, and

speech. Among these are tree diagrams, poetic “feet,” and metrical grids, all capturing the potential for these modes of expression to suggest multiple layers of accent.⁽¹⁰⁾ Examining more qualitative criteria, authors have shown that text painting and text setting are fruitful subjects for in-depth analysis of expression, genre, individual composers’ style, and narrative.⁽¹¹⁾

[1.12] Throughout the vast and growing literature above on texted music and music-speech intersections, ranging from music theory and analysis to cognitive science and linguistics, the assumption is that the linguistic features of the lyrics (e.g., syllabic stress) either align with the way the lyrics are sung (e.g., phenomenal accent in or affecting the melody)—or that they should.⁽¹²⁾ However, not only *can* text and music conflict, but they do so in a variety of ways, to varying degrees, and with different effects on our listening experience.⁽¹³⁾ I define *prosodic dissonance* as any conflict between a text’s prosodic linguistic features and its musical rendition. This could include misalignment between syllabic and melodic stress, between spoken and sung phrases, or between spoken and sung intonation.⁽¹⁴⁾ Note that melodic stress refers to phenomenal accent rather than metric accent, meaning that prosodic dissonance is a distinct phenomenon from syncopation. Whereas syncopation describes the relationship between the meter and a given musical stream (e.g., the melody), prosodic dissonance describes the relationship between the melody and the text. For example, in Example 1, the syllable “-tion-” is not syncopated, but it is dissonant. I will explore some intersections and differences between syncopation and prosodic dissonance further below.

[1.13] The remainder of this article comprises five parts. Part 1 presents the spectrum of prosodic dissonance to consonance in terms of *pitch* and *timing*. Part 2 delves into the language elements (prosody), exploring English linguistic features such as vowel reduction and rhyme as they interact with sung melody. Part 3 returns to the musical sphere, specifically parallelism, syncopation, and phrase breaks. Combining these elements, Part 4 presents three analytical vignettes from Kesha, Royal and the Serpent, and Rina Sawayama. Part 5 concludes by positing how future research might expand or incorporate the concept of prosodic dissonance.

2. Definitions and Parameters

[2.1] If prosodic dissonance is a mismatch between linguistic and musical features, then the more these features conflict with each other, the stronger the dissonance. Similarly, the more they align, the stronger the consonance. Some aspects of language and music lend themselves more easily to such an alignment or misalignment. English has some features that music lacks (at least explicitly), such as lexical categories, semantic content, and consonants/vowels.⁽¹⁵⁾ Similarly, music has some components that language does not explicitly, such as multiple simultaneous streams and a high degree of repetition. What they *do* have explicitly in common is stress—the most natural avenue for prosodic dissonance in English. We might imagine a stress discrepancy space as shown in **Example 1**,⁽¹⁶⁾ where for any given moment, the more stress in the music (particularly the melody) compared to the less stress within the corresponding lyric, the more prosodically dissonant that moment will sound. Inversely, when the relative stress or accent levels between music and text align, the more prosodically consonant it will sound.

[2.2] In English, stress can occur at the level of an individual word, such as “unconditionally,” or at a phrase level, such as “I’m reading a book about music.” The default stress of a spoken phrase adds an additional layer, particularly given that phrasal stress is the only way of accounting for prosodic dissonance on monosyllabic words. Nevertheless, word stress has more potential for dissonance: For multisyllabic words, an unusually stressed word upends the prosody of the entire phrase (“I’m reaDING a book About music”), while increasing the relative emphasis of a word within the phrase can in some cases simply change the meaning (e.g., “I’m reading a book about music; *she’s* reading something else”).⁽¹⁷⁾ For this reason, word stress is usually involved in the sense of misalignment, but particularly uncommon or semantically irrelevant phrase emphasis can also evoke prosodic dissonance. The less common a word’s default stress within a phrase (e.g., “I have an idea”), the more likely it becomes that emphasizing it melodically would sound dissonant.

[2.3] At this point, one major issue emerges: pronunciation, including prosody, depends on language, dialect, and individual speech expression. Syllable stress cannot be absolutely quantified

without flattening the variety and identity attached to language. Though there are “official” recognized pronunciations for individual words, available in dictionaries, textbooks, and public educational resources, it bears remembering that these accounts capture (at best) the pronunciation of the majority of speakers and (at worst) the pronunciation in the dialect(s) of those in power. There are overall trends in pronunciation at population levels that make the discussion still worthwhile—or mutual intelligibility would be more of an obstacle than it is—but keeping that in mind, I do not advocate for strictly quantifying prosodic dissonance. Rather, I prefer to frame the degree of dissonance or consonance comparatively, relative to the surrounding context—with the understanding that the analyst’s own speech will be reflected in their interpretations.⁽¹⁸⁾

[2.4] Regardless, despite variations on the stress of individual words, stress is typically easily audible and intuitive for native speakers of English. In particular, the relationship between sung and spoken stress will likely be felt in relation to two main elements present in both: pitch and timing.

Pitch

[2.5] In English, one way of indicating spoken stress is by placing the stressed syllable on a higher (or, less commonly, lower) pitch than its surrounding contexts. In short, pitch changes mark distinctions. For instance, given wordless sentences with intonation only, as in **Audio Example 2**, it would be possible to determine which syllable is stressed, or even to match up the wordless sentences with text. Similarly, in music, the highest (or, less commonly, lowest) note of a phrase lends itself more easily to perception of accent, particularly if approached by leap.⁽¹⁹⁾ For a prosodically consonant setting, then, we would expect the highest or most distinctive pitches in the music to align with the highest or most distinctive pitches in the text as we’d expect it to be spoken.

[2.6] Examples 2–4 are based on the first line of Taylor Swift’s “You Need to Calm Down” (2019). If we consider the melody without the text, we could compose a prosodically consonant set of lyrics, such as “going for a run in the morning” (**Example 2**). Similarly, if we consider the text alone, we could compose a prosodically consonant melody such as the one in **Example 3**, which emphasizes the stressed syllables “you,” “some-,” “don’t,” and “know.” In Taylor Swift’s rendition (**Example 4**), the single higher pitch in a stream of repeated pitches produces the effect of emphasis on “-dy.”⁽²⁰⁾ Here and for the remainder of the article, yellow circles show musical stress, while purple rectangles show stressed syllables.

Timing

[2.7] Coincidentally (or perhaps not), the highest pitch in the preceding phrase is also the longest, compounding the sense of accent. As discussed above, English has been described as a stress-timed language; its speakers expect stressed syllables to last longer on average than unstressed syllables. When sung, the lengths of syllables may be longer, with more flexible lengths. However, we still tend to hear longer notes, or longer inter-onset intervals (IOIs), as more stressed, and shorter notes or IOIs as weaker.⁽²¹⁾ Furthermore, unlike conversational speech, music adds a metric component, which can compound or conflict with the effect of pitch. (I will discuss intersections with syncopation below in Part 3.)

[2.8] In another one of Taylor Swift’s songs, “...Ready For It?”(2017), the bridge provides another example of durational prosodic dissonance, as shown in **Example 5**. The repeated lyric “Baby, let the games begin” would be consonant with a musical setting placing “ba-,” “games,” and “-gin” on the longest (and if unsyncopated, metrically strongest) notes. Instead, Taylor Swift sings the syllables completely evenly, pausing only after “-gin.” The syllables “ba-” and “-gin” are consonant here, but the equal duration and (unsyncopated) metric placement facilitate hearing “the” rather than “games” as the accented syllable, creating a moment of dissonance. In addition to “the” occurring on the beat, what furthers this effect is not so much a *lengthening* of “the” as a relative *shortening* of “games.” If we recompose the excerpt with a longer duration of this syllable, as in **Example 6**, the dissonance would be eliminated.

[2.9] “. . .Ready For It?” also falls into a broader paradigm of prosodic dissonance that I see emerging in recent years: an evenly spaced melodic rhythm paired with an irregular rhythm in the lyrics. I call this *prosodically dissonant isochrony*, referring to the fact that the equal subdivisions of the melody (isochrony in the music-theoretical sense) conflict with the “equal” spacing of stressed syllables (isochrony in the linguistic sense). To illustrate this paradigm, compare the following excerpts from Dua Lipa, “Levitating” (2020) and Olivia Rodrigo, “good 4 u”(2021) (**Examples 7 and 8**). I encourage the reader to speak the text to each example aloud before and after listening to the musical renditions. Dua Lipa’s verse aligns the “isochronies” of the text and the melody, placing “had,” “-ni-,” “fell,” “rhy-,” and “mu-” on the pitch changes (and on the beats, with no syncopation) and lengthening the longer words “don’t” and “stop.” By contrast, the stressed syllables “car” and “-reer” land on less musically stressed moments of Olivia Rodrigo’s phrase: the pitch accents, which reinforce rather than undermine the beats, seem to emphasize the unstressed syllable “ca-.” As with Example 5, the even spacing seems to “shorten” the syllables that would be lengthened in speech. Though Examples 7 and 8 both feature evenly subdivided melodies landing squarely on the beats, the rhythmic experiences of the two differ enormously through the lyrics. Prosodically dissonant isochrony provides a uniquely clear illustration of how our rhythmic experience changes with our linguistic expectations.⁽²²⁾

Prosodic Dissonance Litmus Test

[2.10] With pitch and timing as our basic parameters, we can begin to build a procedure for identifying prosodic dissonance.

1. Determine the prosody for the lyrics as you’d expect them to be spoken, ideally by speaking them aloud. Which syllables are strongest (most pitch-accented; longest) and weakest (least pitch-accented; shortest)?
2. Determine the prosody for the melody as sung, ideally by singing it aloud. Where are the phenomenal accents (particularly registral and durational)? If the moment in question is not syncopated, where are the strongest metric positions?

Given the first two steps, identify and interpret prosodic dissonance:

1. Identify places where the most stressed spoken syllables match up with the most unstressed sung notes or vice versa. These are moments of prosodic dissonance. The stronger the contrast, the stronger the dissonance.
2. If the moment in question does not sound prosodically dissonant, consider: Is there an alternate pronunciation of the text (e.g., contrastive stress, multiple accepted pronunciations), or an alternate hearing of the melody (e.g., different interpretation of the accent structure), that can account for the unexpected consonance?

[2.11] To demonstrate this method, we can apply our litmus test to our initial Katy Perry example (transcribed in **Example 9**). The word “unconditionally” places stress on “un-” and “-di-” and the least stress on “-con-” and “-tion-”; the melody places the highest, longest, and (given a lack of syncopation) metrically strongest note on “-tion-.” In other words, the “mispronunciation” effect comes from an interaction of pitch and timing that produces strong prosodic dissonance. With the definitions and variables in place, we can proceed to examine some intersections and interactions between prosodic dissonance and linguistic and musical features.

3. Linguistic Features

Vowel Reduction: Accented and Created Schwas

[3.1] In English, stress strongly depends on the phenomenon of vowel reduction.⁽²³⁾ Weaker syllables occur more quickly to reach the next stressed, long syllable. In this process, vowels on weak syllables become more centralized in the mouth, most often becoming schwas.⁽²⁴⁾

Conversely, schwas are associated with weaker syllables. In fact, this may be the most aurally salient marker of stress/unstress in English, and intensifies difficulty understanding familiar words with shifted stress.⁽²⁵⁾ These sonic signals of “weakness” open up at least two possibilities for creating and/or intensifying the effect of prosodic dissonance: accenting a typically reduced vowel or creating a new schwa from a syllable not typically reduced.

[3.2] As an example of an accented schwa, see the second verse to Lady Gaga, “Applause,”(2013) transcribed in **Example 10**. In Lady Gaga’s English, the first and third syllables of the word “nostalgia” contain reduced vowels, with only the middle syllable emphasized: “nə-STAL-јə.” In her sung rendition, however, the pitch accent on the third beat lands on an unstressed syllable, creating prosodic dissonance. I argue that this dissonance is at least more noticeable, if not further intensified, by the melodic accent’s coincidence with a schwa.

[3.3] In the opposite case, some English-language singers reduce the vowel on a *stressed* syllable in prosodically dissonant contexts. Not only does this indicate that the singer also likely experiences this moment as prosodically dissonant, but it further contributes to the melody’s effect on the internal rhythm of the text for the listener. The prechorus to 5 Seconds of Summer, “Easier”(2019) includes one moment where this occurs (**Example 11**). At multiple points, Luke Hemmings reduces the vowel in the word “damn,” which would (particularly being an expletive) typically be emphasized in speech. This “created schwa” effect calls additional attention to the prosodic dissonance, compared to the timing alone.

Rhyme Creation and Emphasis

[3.4] Prosodic dissonance often appears in relation to rhyming words and syllables. In this case, dissonance may seem to create a rhyme in the music that does not exist—or that exists in a much more subtle or understated fashion—in the text as it would be spoken. With the placement of musical (usually melodic) emphasis on typically unstressed syllables, weaker syllables and even reduced vowels become available for rhyming for expressive purposes.⁽²⁶⁾ This phenomenon is well documented in rap, where rappers sometimes use pitch, timing, or vowels to emphasize rhymes and create rhythmic variety.⁽²⁷⁾ However, it also occurs outside of rap, prosodically consonant contexts, and syncopated contexts as well. **Example 12** shows an excerpt from a verse of Janelle Monáe, “Make Me Feel” (2018). Here, Janelle Monáe emphasizes the first syllables of “compression” and “confession” with pitch and timing, creating a rhyme between the weaker, typically vowel-reduced syllables “com-” and “con-.”⁽²⁸⁾ The words’ stressed syllables already rhyme, but that rhyme becomes secondary in the melodic context.

[3.5] While the previous example seems to create a rhyme that would otherwise be absent, prosodic dissonance can also strengthen weak internal rhymes at a phrase level. An excerpt from the song “Stay”(2021) by The Kid LAROI and Justin Bieber, reproduced in **Example 13**, uses strong pitch and timing indicators in the melody to shift the primary rhyme effect from “would/could” to “same/change/can’t/stay.” Though none of these vowels are reduced, only “change” and “stay” would receive stress in a spoken context. The repetition of the same melodic gesture reinforces the other two rhyming syllables, “same” and “can’t”—in other words, using parallelism to emphasize less prominent rhymes.⁽²⁹⁾

[3.6] These linguistic features can be incorporated into our litmus test for additional nuance as follows (additions in underlined bold):

1. Determine the prosody for the lyrics as you’d expect them to be spoken, ideally by speaking them aloud. Which syllables are strongest (most pitch-accented; longest; with the clearest vowels) and weakest (least pitch-accented; shortest; vowel-reduced)?
2. Determine the prosody for the melody as sung, ideally by singing it aloud. Where are the phenomenal accents (particularly registral and durational)? If the moment in question is not syncopated, where are the strongest metric positions?

Given the first two steps, identify and interpret prosodic dissonance:

1. Identify places where the most stressed spoken syllables match up with the most unstressed sung notes or vice versa. These are moments of prosodic dissonance. The stronger the contrast, the stronger the dissonance.
2. If the moment in question does not sound prosodically dissonant, consider: Is there an alternate pronunciation of the text (e.g., contrastive stress, multiple accepted pronunciations), or an alternate hearing of the melody (e.g., different interpretation of the accent structure), that can account for the unexpected consonance?
3. **Consider how the prosody changes the hearing of the text. Does it create or emphasize aspects of the lyrics (e.g., rhyme)?**

4. Musical Features

Parallelism

[4.1] In Example 13 above, the rhyme is emphasized not only by pitch and timing within each individual line, but also by the repetition of the melody. The establishment of sameness and difference makes a strong impact on listener expectations through the effect of parallelism: If the pitches, rhythms, and duration repeat, we expect the prosody of the text to repeat as well. If the prosody changes while the melody does not, this has the potential to create prosodic dissonance through parallelism. Prosodic dissonance increases in salience and intensity when a stress pattern for a melody has already been established only to be changed on a subsequent reiteration. For highly repetitive melodies, this can also serve as a means for musical artists to add rhythmic variety through the text, even while maintaining consistency in other elements of the music.

[4.2] **Example 14** transcribes an excerpt from Ava Max, “My Head and My Heart”(2020), which follows exactly the trajectory described above. The melody, including its stress pattern, remains the same through the first two and a half lines. The text also remains relatively constant in prosody: “I think about me now and who I could’ve been / And then I picture all the perfect that we lived.” Without the text, the third line sounds as an exact repetition of the first two; however, the lyrics change emphasis: “‘til I cut the strings on your tiny violin.” The lyrics’ emphasis moves from the fourth syllable (“the”) to the fifth (“strings”), creating prosodic dissonance through contrast with the previous lines and through the high-pitch, downbeat placement of “the.” One could imagine a recomposition of this passage to place the pitch accent on “strings,” creating a syncopation effect instead of prosodic dissonance (**Example 15**). Even then, the effect of parallelism creates fertile ground for dissonance, particularly in comparison to the consonant end of the line, “tiny violin.”

Interactions with and Differences from Syncopation

[4.3] Previous scholars have documented a separate but related phenomenon regarding prosody and music: the use of syllable stress to create or reinforce syncopation (Temperley 1999, VanderStel 2021).⁽³⁰⁾ In both (stress-based) prosodic dissonance and syncopation, stress patterns misalign. However, as mentioned above, syncopation involves one or more musical layers (for our purposes, the text) misaligning with the stress pattern of the *meter*, while in prosodic dissonance, the stress pattern of the text (prosody) is dissonant with the stress pattern of the *melody* (typically defined by patterns of durational and registral accents)—whether it aligns with the meter or not. Prosodic dissonance may occur between the text and metric accents, but syncopation has the potential to undermine that effect. For example, the verses to Billie Eilish’s “bad guy” (2019) frequently anticipate the meter with the melody (**Example 16**), with many of the notes shifted off the beat. Despite the shifting, this doesn’t feel particularly dissonant: “like,” a strong syllable, gets a longer duration, as does the stronger syllable in “control.” The line “I told you I’d change” in Example 13 above also matches syllabic and melodic stress that does not land on the beat. Because the two phenomena involve different types of interactions between musical layers (text/metric accent vs. text/melodic accent), syncopation and prosodic dissonance can interact or operate independently. A moment may be syncopated but prosodically consonant, it may be prosodically dissonant *and* syncopated, it may be prosodically dissonant without syncopation, or, of course, it may be

prosodically consonant and unsyncopated. **Examples 17–20** provide hypothetical, simple examples of each.

[4.4] In Example 17 an even, unsyncopated quarter-note pace pairs with four consecutive iambs (or weak-STRONG pairs): “mail,” “-rived,” “three,” and “clock.” Although the rhythm changes from the spoken rendition, which would shorten the weaker syllables compared to the stronger ones, the relative emphasis of the lyrical and musical accent streams still align, making the excerpt prosodically consonant. Contrast this with Example 18, which takes the accented syllables “-view,” “-vise,” “don’t,” and “pa-” and places them musically in various states of stress. Yet, the accent pattern of the melody does not allow for a syncopated hearing; it is unclear which events would be stressed to allow such an interpretation. Example 19 provides the opposite context, with the stressed syllables of the text evoking a *tresillo*.⁽³¹⁾ Because of this pitch accent and apparent syncopation, “de-” and “-vered” do not sound prosodically dissonant, particularly if we reduce the vowel for the first syllable. Finally, Example 20 uses both prosodic dissonance and syncopation: syncopation in the melodic stream (which, like Example 19, is grouped by contour into 3+3+2), and prosodic dissonance in the resulting emphasis of “de-” and “-ry” as well as a relative shortening of “lunch.”

[4.5] Given that word/sentence stress can specifically lead us to hear syncopation, the two effects can sometimes be difficult to differentiate. In these cases, a prosodically dissonant hearing and a syncopated hearing can be framed as competing modes of listening: If the linguistic features (such as syllabic stress patterns) are prioritized, or seem to overwhelm the musical features, the lyrics can create or reinforce syncopation—in other words, the features of the language are superimposed onto the music. If, in contrast, we mentally prioritize the musical features, this can overwrite the features of the lyrics, causing us to hear prosodic dissonance.

[4.6] The ambiguity between syncopation and prosodic dissonance is particularly salient when we move beyond the accent structure of the melody to other musical layers, as in **Example 21** (another excerpt from “Applause”). The moment in question is the sung rendition of “art’s in pop culture in me” at the end of the verse. The stressed syllables here as spoken are “art,” “pop,” “cul-,” and “me.” Without any clear syncopation in the melody, and with a repeated step down gesture on each of the four quarter notes, the emphasis ends up on “-ture” instead of “cul-,” causing a moment of dissonance. Parallelism intensifies this further, with the previous phrases ending with “some of us just like to read” and “suddenly the Koons is me,” both of which have a prosodically consonant strong syllable on beat 3.⁽³²⁾ However, drawing our attention to the accompanying tracks rather than the melody alone can lessen the experience of prosodic dissonance and even shift it to a syncopated hearing. The melodic synthesizer present from the very beginning of the song, transcribed in **Example 22**, articulates a *tresillo* pattern behind each phrase, which conflicts with the first two lines of the lyrics but comes to fruition with the similarly *tresillo*-evocative prosody of “art’s in pop culture in me.” While this article has so far focused exclusively on melody, since the singing itself is a natural place to look for prosodic dissonance, this example demonstrates that other musical layers can also affect the prosody, if less overtly. By shifting our focus to different aspects of the music, we can toggle between two distinct rhythmic listening experiences of this passage—parallelism or the synthesizer, speech or song, syncopation or prosodic dissonance.

[4.7] In cases like “Applause,” we may be able to control or choose our hearing by shifting our focus, but there are elements that contribute more generally to the possibility of hearing something as syncopated or prosodically dissonant. The more musical emphasis on a weak syllable—through pitch, duration, metric position, timbre, articulation, or a combination of the above—the more prosodically dissonant a moment will sound. Similarly, the less musical differentiation or accent there is—or the more musical support for a strong syllable there is—the more possible it becomes to hear syncopation. Parallelism, vowel shape, and rhyme, discussed above, can further reinforce one of these hearings. For example, if we removed the octave leap in Example 19 and lengthened the weak syllable, the resulting musical emphasis on “-vered” would tip the balance back toward prosodic dissonance, as shown in **Example 23**. Yet if we were to change the setting to be as uniform as possible, as in **Example 24**, we would likely still favor a syncopated hearing. In other words, *when possible*, we prefer a hearing that aligns features of music and lyrics.⁽³³⁾ Since rap frequently

uses only minimal melodic features, this preference also accounts for the relative lack of prosodic dissonance—and relative prevalence of lyric syncopation—in rap music as compared to sung music.⁽³⁴⁾

[4.8] The availability of a syncopated hearing also connects to the accent structure of the text itself. If stressed syllables in the lyrics are more evenly distributed, we may be able to mentally shift our musical accent structure to hear syncopation. If, however, the text has adjacent stressed syllables (sometimes called a spondee in discussions of poetic rhythm) or long strings of weak syllables, it may be much more difficult to conjure a syncopated version of the text. We can speak with adjacent stressed syllables in English, but we cannot perceive two quick adjacent accents in music as easily (Ohriner 2019b). In Example 18, the three adjacent accents in “-vise, don’t pa-” render it more difficult to “shift” our sense of accent to hear it as syncopated. In short, features of the music and language can privilege or exclude prosodic dissonance, privilege or exclude syncopation, or contribute to a paradoxical experience where both are simultaneously present.

Phrase Breaks

[4.9] Thus far, all the features discussed linguistically and musically have centered on stress and accent. However, prosodic dissonance consists of *any* conflict between linguistic and musical features. Another related but distinct feature of music and language is the concept of *phrase*, particularly boundaries between phrases, marked most frequently by pauses (timing). Particularly since streams of spoken language do not usually place pauses between individual words, what pauses and breaths are present usually indicate breaks between ideas and thoughts. Similarly, we teach our students to listen for resting points to distinguish musical phrases.⁽³⁵⁾

[4.10] When the phrases in the music and lyrics begin and end at different times, this produces a distinct category of prosodic dissonance, which I call *enjambment dissonance*.⁽³⁶⁾ More specifically, enjambment dissonance occurs either when a musical phrase continues across what would constitute a phrase boundary in speech, creating a run-on effect, or when a musical phrase boundary occurs in the middle of a spoken phrase, creating an interruption effect.

[4.11] For an example of the former, see **Example 25** from Frankie Bird, “Paper Doll”(2017). The song begins with an ambiguous case of prosodically dissonant isochrony, creating an effect somewhere between syncopation and prosodic dissonance as she sings “wake up in the morning.” The following lyric would normally be composed of two units, with a brief lengthening of the duration between them: “got no makeup on / that’s fine by me.” However, Frankie Bird’s rapid, even delivery eliminates the space between the two lines of text, creating enjambment dissonance on top of the prosodically dissonant isochrony of the individual phrases. Compare this sound to **Example 26**, an excerpt from Ariana Grande, “positions”(2020). The lyrics to this passage, “I’m just hoping I don’t repeat history,” would typically be spoken as one unit. Ariana Grande’s delivery separates them into two segments, adding space between the two syllables of “repeat” to create a feeling of interruption. Furthermore, her choice to pronounce rather than reduce the vowel in “re-” further reinforces the phrase-ending emphasis on this syllable.⁽³⁷⁾

[4.12] As these two examples show, enjambment dissonance can co-occur and interact with syllable-stress-based prosodic dissonance. The elements in common between music and language offer more than just opportunities for text painting, musical speech or speech-like music—they highlight a fundamental difference between texted and untexted music. Text and music can constitute two distinct layers of sound: layers whose independence is imperceptible without illuminative conflict between them. Such conflict—prosodic dissonance—is rife with analytical and interpretive questions.

[4.13] Before applying these principles to some more extended musical examples, it bears updating our procedure one final time:

1. Determine the prosody for the lyrics as spoken, ideally by speaking them aloud.
Which syllables are strongest (most pitch-accented; longest; with the clearest vowels) and weakest (least pitch-accented; shortest; vowel-reduced)? **Where do we put pauses**

or resting points? Is there any parallelism with the prosody of the surrounding lyrics?

2. Determine the prosody for the melody as sung, ideally by singing it aloud. Where are the phenomenal accents (particularly durational and registral)? If the moment in question is not syncopated, where are the strongest metric positions? Where do we hear pauses or resting points? Is there any parallelism with the prosody of the surrounding melody? Does the melody, or one of the other musical layers, lend itself to hearing syncopation that changes any of the above interpretation?

Given the first two steps, identify and interpret prosodic dissonance:

- a. Identify places where the most stressed spoken syllables match up with the most unstressed sung notes or vice versa.
- b. Identify places where the most space between spoken syllables matches up with the least space between sung notes or vice versa (enjambment).

These are moments of prosodic dissonance. The stronger the contrast, the stronger the dissonance.

3. Consider how previous material shapes our hearing of prosody. Does any parallelism intensify or lessen the apparent misalignment?
4. If the moment in question does not sound prosodically dissonant, consider: Is there an alternate pronunciation of the text (e.g., contrastive stress, multiple accepted pronunciations), or an alternate hearing of the melody (e.g., syncopation, different interpretation of the accent structure, parallelism with a previous event), that can account for the unexpected consonance?
5. Consider how the prosody changes the hearing of the text. Does it create or emphasize aspects of the lyrics (e.g., rhyme)?

5. Analysis and Tension

[5.1] To showcase this more nuanced version of this process, I offer three brief interpretive vignettes featuring prosodic dissonance in different ways.

Variety and Destabilization in Kesha, “Tonight”

[5.2] Kesha’s 2020 song “Tonight” depicts an energetic, drug-filled night of partying. In this song, Kesha alternates between the powerful and emotional pop-ballad vocal persona of her more recent songs (e.g., “Praying”) and the pitch-expressive, creaky-voiced rapping characteristic of her earlier hits (e.g., “TiK ToK”).⁽³⁸⁾ The verses embody the latter persona, each consisting of four two-measure streams of even subdivisions.

*Okay, we’re going out tonight, don’t wanna stay home
I got my girls to call the Uber ‘cause I can’t find my phone
I’m getting ready, mani-pedi, fancy shit with the leathers
Now we’re looking for some trouble like we hunting for treasure*

[5.3] These four lines group into two pairs, with each pair having matching rhythms and rhymes. The second pair of lyrics contrasts with the first by adding a syllable and changing the rhyme. However, the distribution of stressed syllables stays the same between the first and second pairs, which is also reflected in Kesha’s musical delivery. “Home” and “phone,” despite occurring ahead of the beat, receive a longer duration such that the melodic accent still aligns with the syllabic stress. In the following two lines, Kesha places a higher relative pitch on the strong syllables “lea-” and “trea-” such that the text is still prosodically consonant. We could represent the contour of this verse, as well as its corresponding stress pattern, as shown in **Example 27**.

[5.4] The second verse's prosody follows a different trajectory. Once again, Kesha raps in even sixteenth-note subdivisions with anticipatory syncopations on the last syllable of each line. This time, however, all four lines have exactly the same rhythm of twelve sixteenth notes, each ending with an [æk(t)] rhyme.⁽³⁹⁾ The first two lines follow the same pattern as the first verse:

*Okay, we staying out tonight, there's no turning back
I got my shorties up so high, bet y'all think I'm running track*

The second couplet, in contrast to the first verse, makes significant changes to the prosody while maintaining the same number of syllables.

*Just found out me and Elton John have the same shoes, that's a fact
Hey Chelsea, do you mind if I put this wine in your backpack?*

[5.5] The stressed syllables “me” and “John” sound consonant in parallelism with the previous line, particularly since they align with the beat (and a slight heightened pitch on “John”). From here, we would expect “same,” “shoes,” and “fact” to be emphasized, with a pause where the comma is. Instead, five elements contribute to a prosodically dissonant setting that beautifully summarizes the above. First, Kesha’s rapping intonation places pitch accents on “the” (a weak syllable) and “shoes,” higher than the intervening “same” (a strong syllable). Second, the text itself has two consecutive stressed syllables (“same shoes”), making a mental shift to syncopation impossible. Third, the emphasis on the vowel-reduced syllable “the,” coupled with a slight backing-up and near monophthong-ization of the diphthong in “same”—[seɪm] becomes something closer to [seɪm]—adds phonetic emphasis to the dissonance. Fourth, the parallelism from the previous two lines’ consonance makes the text stick out by comparison. Finally, the short duration between “shoes” and “that’s” (where we would expect a pause) creates enjambment dissonance before bringing the listener abruptly out of the prosodic dissonance and back into the following consonance (“that’s a fact”). Within a few seconds, this third line of text begins with prosodic consonance (“Just found out me and Elton John”), abruptly shifts to dissonance (“have the same shoes”), and ends with brief consonance (“that’s a fact”). In contrast with the previous verse and two preceding lines, this drastically upends Kesha’s established prosodic flow, introducing instability and ambiguity.

[5.6] The next line of lyrics, “Hey Chelsea, do you mind if I put this wine in your backpack?” achieves a different effect from any of the previous three lines. If we listen according to the prosodic rhythm established by the first two lines—the same rhythm that contributed to the dissonance of the previous line—this lyric would also begin from a point of relative consonance (“hey Chelsea, do you mind”) and move to dissonance (“if I put this wine in your backpack”). However, Kesha’s delivery pitches “I,” “wine,” and “back-” slightly higher relative to the surrounding syllables, supporting a prosodically consonant, syncopated hearing with a *tresillo* starting on beat 2. This supports and is supported by the internal rhyme “mind” and “wine.” The dissonance in the third line upends the established prosodic rhythm, providing a transitory state to move to a consonant, but more syncopated, flow.⁽⁴⁰⁾ Example 28 shows a transcription of the second verse.

[5.7] The interaction of prosody and syncopation here adds variety and direction to four otherwise rhythmically identical lines of lyrics with the same rhyme. At this point, we could add an interpretive “step 7” to the litmus test. In this case, I would suggest that the messy and fluid transition between prosodic states adds to the narrative of the song, effecting a loose, clumsy, spontaneous sound to match the drunken, carefree attitude of Kesha’s narrator. As the verse progresses, the text moves from rigid, consistent, and organized to loose and flexible, just as the narrator moves from sobriety to inebriation. The narrative of a wild night out extends to the delivery of the lyrics themselves.

Expressive Enjambment and Text Painting in Royal & the Serpent, “Overwhelmed”

[5.8] Royal & the Serpent’s song “Overwhelmed,” also released in 2020, makes pervasive use of prosodic dissonance, including enjambment dissonance and rhyme interactions. This relationship

begins right from the start of the song, with a stark discrepancy between the musical delivery and groupings of the lyrics (**Example 29**).⁽⁴¹⁾

[5.9] The first few words, “Turn off the TV,” already sound slightly dissonant on their own; the relative high pitch of “T-” suggests a syncopated hearing that emphasizes this syllable (whereas “TV” would often be pronounced with two consecutive stressed syllables). The next lyric, however, overwhelms this sense of dissonance with a much more noticeable instance: “it’s starting to freak me—.” Royal & the Serpent breaks off the musical phrase in the middle of the text’s phrase, using silence to reinforce the effect, and then continues as the next musical idea begins, omitting the break between “out” and “it’s.” Parallelism from the first line (“starting to freak me . . .”) also contributes to hearing an accent on “my” rather than the spoken expectation “ears.” A better representation of the musical delivery, then, would be **Example 30**.

[5.10] This visual representation of enjambment dissonance in the lyrics, altered to reflect the *musical* context rather than the structure of the lyrics alone, is also used by all three of the top “lyrics videos” for the song on YouTube at this time of writing, indicating that the perception of this enjambment dissonance is widely heard by listeners.⁽⁴²⁾ The lyrics and the music both use a short-short-long motivic structure, but whereas the melody uses two short ideas of equal length, the text’s short ideas have different numbers of syllables, with phrases gradually increasing in length. The notion of “short-short-long” therefore effects different rhythmic experiences as we hear conflicting phrase rhythms through the prosody. In this way, Royal & the Serpent outlines a metaphorical arch shape from accent-based prosodic dissonance to enjambment dissonance and back again, affecting the perception of tension in the first two lines. Setting a precedent that will recur throughout the song, the enjambment-dissonant placement of “out” also reinforces the internal rhyme with “loud.”

[5.11] The following two lines, despite having the same melody and rhythm, provide contrast with relative prosodic consonance in the places where dissonance occurred previously (**Example 31**). The first verse leading up to the prechorus, then, follows a path from dissonant to consonant phrase breaks, creating tension and release between the first and second couplets. Before discussing the prechorus and chorus sections, let us compare this verse to the second verse, which heightens and develops some of these effects.

[5.12] Liaising with the title line that ends the chorus, the second verse also begins with prosodic consonance: “All of these faces who don’t know what space is.”⁽⁴³⁾ From here, Royal & the Serpent’s prosody begins to desynchronize with the melody once more, culminating with an apparent delivery of “overstimulated” (rather than “overstimulated”), a hearing supported by previous parallelisms as well as the recurring melodic accents on F. I interpret this prosodic dissonance as a form of text painting, illustrating the narrator’s discomfort, sense of nonbelonging, and difficulty conforming to expected social norms. This process continues in the next line, which once again begins with consonance (“nobody gets it”) before placing the strongest dissonance on a word describing the narrator (“sensitive”). The placement of “-tive” in the following musical subphrase creates the same enjambment dissonance pattern as the very first couplet, this time breaking a musical phrase in the middle of a word. In the line following “sensitive,” Royal & the Serpent adds more syllables, creating the most syllable-dense line of both her verses together. The setting is relatively consonant, stressing the strong syllables “list-,” “eye-,” and “ex-” as shown in **Example 32**; however, the inserted additional syllable contributes to the information density, further expressing the sentiment of the song.

[5.13] In both verses, particularly the second, prosodic dissonance reinforces the narrative, conveying the speaker’s feelings of overstimulation and sensitivity. The sense of social rigidity or expectation is further supported by the insistent repetition of the pitch F as every single melodic accent. Interestingly, this rigidity does *not* always match the meter; anticipatory syncopation is plentiful throughout this song. At some points, the prosody aligns with this syncopation; in other places, prosodic dissonance arises as an effect of ignoring this syncopation. The awkwardness and disjointedness of some of the lyrics, coupled with the rigidity of the metronomic texture, make the lyrics *themselves* sound overwhelmed and out of place, as if the speaker’s emotions are struggling to fit into the confines of the song and conveying the sentiment in the song’s lyrics.

[5.14] The prechoruses and choruses are more consistently consonant by comparison, suggesting a prosodic variant on the “loose verse, tight chorus” model (Temperley 2007)—or textual stratification, if we consider text to be a textural element (Covach 2018). While the verses play with strong enjambment and accent dissonances to convey discomfort, the percussion and repetition in the chorus express the narrator’s anxiety through a rigid, inflexible melodic accent structure. Finally, one additional moment of timing-based prosodic dissonance in the chorus further contributes to the text painting: Royal & the Serpent’s delivery of the line “makes it hard to breathe” slightly condenses the expected pacing of these syllables, placing “hard” and “breathe” closer together compared to “makes.” This suggests a rushing through the syllables, directly conveying the breathlessness expressed in the lyrics themselves.

Shifting Parallelism and Rhyme in Rina Sawayama, “This Hell”

[5.15] The first verse of Rina Sawayama’s “This Hell”(2022) establishes the song with a high degree of prosodic dissonance before a largely consonant chorus. Unlike the previous two examples, however, the effect of the prosodic dissonance resists parallelism and recontextualizes lyric emphasis rather than evoking clear text painting.

[5.16] The opening line conveys persistent dissonance throughout as Rina sings, “Saw a poster on the corner opposite the motel.” Whereas the strongest syllables in speech would be “post-,” “cor-,” “op-,” and “-tel,” the musical context suggests an alternative distribution of stress. First, the rhythmic repetition in the melody groups “poster on” with “corner op-,” deemphasizing the latter syllable in favor of the weaker “-site.” The two slightly higher pitches on “-site” and “mo-” then suggest stress for these syllables.

[5.17] The next line, “Turns out I’m going to hell,” reshapes the dissonance on “motel.” With a rhyme and a stressed syllable in the same (final) position of each line, and with a longer duration between the onsets of “-tel” and “turns,” the slight pitch accents on C are overwhelmed by the possibility of a consonant, syncopated hearing, as shown in **Example 33**. The third line’s rhyme, “fl keep on being myself,” further intensifies the shift, to the point where any dissonance on “motel” vanishes in retrospect. Parallelism and rhyme here lead us back to prosodic consonance—as opposed to a hypothetical alternative set of lyrics with a stress pattern reinforcing the initial dissonance and removing the rhyme, such as “although I’m doing nothing” / “they say I’m sinning daily.”

[5.18] Just as this consonance solidifies, the second half of the verse undermines it by beginning abruptly on the downbeat, truncating the length of the stressed syllable “-self” (**Example 34**). This enjambment dissonance is particularly salient in combination with the consecutive stressed syllables “-self” and “don’t.” Softening this effect, the official music video for the song also places a sudden cut between the two syllables, even though the previous three lines are all lip-synced in a single shot. The resulting rapid and difficult consonant cluster “lfd” contributes to a feeling of “self-interruption,” particularly audible in live performances.⁽⁴⁴⁾

[5.19] The following line, “Don’t know what I did,” returns to prosodic consonance momentarily, with the longest pitches placed on the longest spoken syllables “don’t,” “know,” and “did.”⁽⁴⁵⁾ This time around, both “don’t” and “know” sound consonantly accented with the even quarter notes and corresponding hits in the backing track, slightly contrasting with the stress pattern of “saw a” in the previous instance.⁽⁴⁶⁾ The consonance is short-lived as the next lyric begins: “they seem pretty mad about it.” As an effect of the pacing that lengthens “they” and “pret-,” Rina reduces the vowel in “seem” to a schwa, creating a disjunction between the lyrics’ and melody’s stress tendencies. Furthermore, whereas the emphasis in the text “mad about it” would be consonant with stress on “mad” and “-bout” (supported by the pitch accents), the parallelism of the previous lyrics, “going to hell” and “being myself,” gives the conflicting impression of emphasis on the last syllable, “it.” The ending rhyme with “did” intensifies the viability of this hearing.

[5.20] Thus far, in under 13 seconds of music, Rina Sawayama fluctuates rapidly between consonant and dissonant states in her delivery, each state intensifying the effect of the other.

Parallelism first changes us to the less dissonant, syncopated hearing in Example 33, which in turn increases the prosodic dissonance of “mad about it” (see Example 34). Rhyme, parallelism, and syncopation constantly vie for auditory priority, contradicting one another and presenting a variety of possible experiences of the verse—all of which will include some degree of prosodic dissonance.

[5.21] While “This Hell” proceeds to a more consonant prechorus, chorus, and second verse, seeming to respond to some of the tensions of this first verse, one more salient moment of dissonance stands out: the phrase “damned for eternity,” with the final syllable “-ty” emphasized through length, pitch, rhyme with the following line, and parallelism with the previous lines. This moment signals a possible expressive correlation of prosodic dissonance in this song: an association with the homophobic perspective of others. The conflicting prosody of the lyrics seems to represent the restrictiveness of societal expectations, with the consonance that follows expressing the comfort of pride and authenticity. Rina’s prosody seems to mock the discomfort of the angry rioters who would damn her for her queerness.

[5.22] Prosodic dissonance plays a different role structurally and expressively in each of these three examples. By recognizing and qualifying it not (or not only) as a mark against artistic value, but as another avenue for expression, conflict, and variety in texted music, we gain an entirely new layer of analytical opportunities: prosody. This layer cannot exist without the musical setting, since the text would have no pitch, timing, or phrasing to conflict with; it similarly cannot exist without text, since prosodic dissonance would be imperceptible and meaningless without an intuition for that language’s pronunciation.

6. Future Study of Prosodic Dissonance

[6.1] Because they are interdisciplinary by definition, investigations of prosodic dissonance lend themselves well to a variety of further research topics, including intersections with structure and form, corpus studies, differences between languages, cognitive studies, questions of compositional process and reception, analysis of text and musical accent alone, stylistic norms, linguistic change, music theory pedagogy, and notational/visual representations.

[6.2] In all three of the analytical examples in Part 4, the majority of prosodic dissonances occur within the verse, suggesting a prosodic link to Temperley’s (2007) “loose verse, tight chorus” phenomenon. Whereas Temperley’s paradigm describes coordination or separation of melody and harmony (the “loose” descriptor applying to the “melodic-harmonic divorce”), further research could be done on prosodic dissonance as analogous to a “text-melody” divorce.⁽⁴⁷⁾ My initial impression is that prosodic dissonance is less common in choruses, marking “Unconditionally” as a notable exception and offering a partial explanation as to why it received more negative attention than other prosodically dissonant songs. Further analyses, either at the level of individual pieces or entire corpora, could shine light on the formal associations of this phenomenon. Regardless of how verses and choruses interact with prosodic dissonance, we could also experiment with prosodic dissonance as form-defining in the sense that it can shape our experiences of tension, rhythmic complexity, and consonance.

[6.3] As discussed in the introduction, there is some precedent for encoding linguistic stress in musical corpora. This type of feature can be extended to include vowel reduction, distinctions between primary and secondary stress, or even the presence, absence, or ambiguity of prosodic dissonance itself. Prosody could be added as an additional layer to existing corpora or used as the basis for new encoding. Corpus studies that include prosodic dissonance will enable a much wider array of large-scale analytical questions, including tracking the historical or stylistic development of prosodic dissonance, connections to particular artists or years, or even the overall presence or absence of prosodic dissonance within popular songs conceived broadly.

[6.4] This article has restricted its purview to English prosody, but each spoken language will have different elements to align or conflict with musical ones. How does prosodic dissonance change for syllable- or mora-timed languages, such as French or Japanese? Languages with different

indicators of syllable stress or phrasing could lead to new types of prosodic dissonance not available or common in English. Tonal languages, such as Cantonese, also open up the possibility of a new dimension of dissonance based on pitch or contour.⁽⁴⁸⁾

[6.5] I have consistently referred to the strength or presence of prosodic dissonance; behavioral and cognitive studies will be necessary to determine its impact. How do different musical contexts for the same text, or different texts for the same musical context, produce, strengthen, or weaken prosodic dissonance? Which types produce negative value judgments and which do not? What circumstances might lend themselves to “prosodic paradoxes” that split a pool of listeners between syncopated and prosodically dissonant interpretations? Are prosodically dissonant lyrics more difficult to remember or reproduce? These and other research questions offer new avenues into the exploration of music-language intersections.

[6.6] Though participant studies can illuminate one category of causes for prosodic dissonance, we might adopt a musicological approach and search for causes in the songwriting process. Can we trace a rise in prosodic dissonance to particular songwriters or songwriting teams? In cases where the melody and lyrics are penned by different authors, is prosodic dissonance more common? Is syllable alignment a common consideration in the music industry? In short, to what degree are these types of effects compositionally intentional and to what degree are they byproducts of other priorities and processes?

[6.7] The organization of the text or music itself could also be a line of investigation. In cases of prosodic dissonance, we can examine patterns of syllabic stress, regularity, and phrasing. I would hypothesize that a text with a more irregular distribution of stressed syllables would correlate with higher instances of prosodic dissonance, as would highly isochronous streams of equal subdivisions in the melody. In cases with spondees, or two consecutive stressed syllables, I would similarly expect that one rather than both would be musically accented. Lastly, we could propose certain types of musical contour that facilitate or preclude the commonality of prosodically dissonant settings.

[6.8] All of the examples used in this article have been pop (or pop-adjacent) music. However, as noted in the introduction, prosodic dissonance occurs in genres from rap to emo to early-twentieth-century French art song. These contexts all differ from one another; while prosodic dissonance in rap may offer more intersections with rhyme creation/emphasis and syncopation (Duinker 2022), many instances in punk and emo music of the last two decades contribute to the DIY aesthetic noted by previous authors (Chiu and Blake 2021). Analyzing prosodic dissonance in classical genres opens a new, if related, set of questions, since many of the texts predate their musical settings rather than being co-created. Prosodic dissonance can be a marker of musical style or individual artistry, helping to define or cross genre boundaries. The perception of its presence or absence, therefore, will likely be affected by the genre and by the typical listening habits of those asked.

[6.9] Continuing the long list of fruitful future research topics, prosodic dissonance also relates to vocal delivery and timbre. Not only might vowel sounds other than reduction be grounds for prosodic dissonance on their own—e.g. re-pronouncing words in order to rhyme them—I also hypothesize that how speechlike or songlike an artist’s delivery sounds can intensify or weaken the effect of prosodic dissonance. For instance, **Audio Example 3** contains an excerpt from Olivia Rodrigo, “brutal,” which evokes a speaking quality. The prosodic dissonance in this example is particularly jarring, owing in large part to her delivery: the combination of her low conversational tone with the apparent shift in emphasis from the musical context suggests the very awkwardness depicted in the lyrics. Similarly, some melodies with more songlike qualities might soften the effect of dissonance—unless the pitch of the melody seems to add to that conflict. The timing of artists’ breaths can also affect the suggestion of phrase breaks, perhaps independently from duration itself (Beaudoin 2024). The growing subfield of timbre studies benefits from the inclusion of interdisciplinary, linguistic notions such as phonetics, dialect studies, and vocal tone.⁽⁴⁹⁾

[6.10] Of course, language constantly shifts and changes as a result of many interacting social factors. When a mispronunciation of a word becomes common to a group of speakers rather than

one individual, it ceases to be a mispronunciation and becomes a marker of style, dialect, or other meanings. Another area for future research would be to choose music performed by speakers of different dialects (British dialects, American dialects, Australian dialects, Singaporean dialects, AAVE, etc.) and compare their treatment of prosody. It also follows that the perception of prosodic dissonance will depend on the rarity of a given pronunciation—what I call the “somebody effect.”⁽⁵⁰⁾ Even if a word’s pronunciation does not change in speech, its expected sung emphasis can migrate, for example, in the common placement of stress on either of the first two syllables of “somebody” (see **Audio Examples 4** and **5**).⁽⁵¹⁾ Further cognitive and linguistic studies would be necessary to understand more fully the extent and conditions of this overexposure effect.

[6.11] One interesting area for future research, alluded to earlier in the article, is the potential for visualization techniques for prosody as set in music. While there is precedent for using visual elements of the written text itself for this purpose, such as CAPitaliZAtion, *italics*, line / breaks, or spacing, these could be expanded, codified, adapted, and/or replaced with other notational technologies, combining musical and linguistic visualization practices.⁽⁵²⁾ For instance, **Example 35** shows one rudimentary way of representing the gulf between a word as pronounced and as sung. The way we choose to represent sonic phenomena influences the way we experience and perceive them, and there are near infinite possibilities for what can be represented and how.⁽⁵³⁾

[6.12] Finally, the pedagogical opportunities of prosodic dissonance will in turn illuminate more about this phenomenon, its perception, and its creation. Students could experiment with writing prosodically dissonant and consonant settings for a particular text—or prosodically dissonant or consonant text for a particular melody. Other assignments might challenge them to adjust existing songs to increase, lessen, or change elements of prosodic dissonance and consonance, to find some examples themselves, to identify it aurally, or to write analytical essays on patterns and roles of prosody in particular songs or artists’ discographies.

[6.13] I conclude this article with a plea for descriptivism: introductory linguistics textbooks begin by highlighting that the goal of this field is not to “correct” language, but to observe it.⁽⁵⁴⁾ I believe music theory could benefit from this shift at every structural level. When we encounter what we initially interpret as bad text setting, or any other type of “mistake,” particularly in popular music, it would serve us better to approach it with curiosity in addition to or instead of arbitration. While one anomaly could be a sign of misunderstanding what musical traits a genre typically employs, or of importing traits from one style to another, repeated instances of the same anomaly make a pattern that changes what the traits are of the style itself.⁽⁵⁵⁾ Rather than—or at least in addition to—laughing at apparent “mispronunciation” or heralding it as a sign of the decline of art (as amateurs and professional theorists have done throughout history), we only stand to gain from setting aside our value judgments and asking *why* this sounds mispronounced and what that does to our musical experience.

[6.14] So what of “Unconditionally”? We could craft a variety of narratives accounting for Katy Perry’s prosodic dissonance in Audio Example 1 and Example 9. Is it a simple case of bad text setting, a byproduct of music and lyrics composed separately, an unintentional use of prosodic dissonance that can nevertheless support our analytical agendas, a reflection of a larger growing discrepancy between English as sung and spoken, or a deliberate “mispronunciation” to gather media attention or make a comment about the conditional love of pop fans? If it is the clarity of compositional intent that determines the difference between artistic expression and bad text setting, we may be missing the point of analysis altogether. There is not, nor can there be, a single “correct” interpretation, but any of the above *could* be supported or at least investigated. Whether or not Katy Perry and her songwriting team intended it, I choose both to laugh uncomfortably at her rendition of the title word, then to interpret my own discomfort exactly in line with the intended message of the song: regardless of apparent flaws, embarrassing missteps, or music-theoretical naïveté, the narrator proclaims acceptance and love of the whole listener.⁽⁵⁶⁾ The song expects the discomfort of our angry tweets and well-meaning pedagogy and dares us to enjoy it unconditionally—whether we ultimately choose to or not.

Eron Smith
Oberlin College & Conservatory
77 W College St, Oberlin, OH 44074
esmith9@oberlin.edu

Works Cited

- Adams, Kyle. 2008. "Aspects of the Music/Text Relationship in Rap." *Music Theory Online* 14 (2). <https://doi.org/10.30535/mto.14.2.3>.
- _____. 2009. "On the Metrical Techniques of Flow in Rap Music." *Music Theory Online* 15 (5). <https://doi.org/10.30535/mto.15.5.1>.
- Albert, Sara W. 2021. "A Theoretical and Experimental Analysis of Prosodic Dissonance in Post-Millennial Pop Music." Bachelor's capstone project, University of Pennsylvania.
- Aroui, Jean-Louis, and Andy Arleo. 2009. *Towards a Typology of Poetic Forms: From Language to Metrics and Beyond*. John Benjamins Publishing. <https://doi.org/10.1075/lfab.2>.
- Arvaniti, Amalia. 2009. "Rhythm, Timing, and the Timing of Rhythm." *Phonetica* 66 (1-2): 46–63. <https://doi.org/10.1159/000208930>.
- Arvaniti, Amalia. 2012. "The Usefulness of Metrics in the Quantification of Speech Rhythm." *Journal of Phonetics* 40 (3): 351–73. <https://doi.org/10.1016/j.wocn.2012.02.003>.
- Attas, Robin. 2011. "Sarah Setting the Terms: Defining Phrase in Popular Music." *Music Theory Online* 17 (3).
- BaileyShea, Matt. 2021. *Lines and Lyrics: An Introduction to Poetry and Song*. Yale University Press.
- Beaudoin, Richard. 2024. *Sounds as They Are: The Unwritten Music in Classical Recordings*. Oxford University Press.
- Bertinetto, Pier Marco. 1989. "Reflections on the Dichotomy 'Stress' vs. 'Syllable-Timing.'" *Revue de phonétique appliquée* 91 (93): 99–130.
- _____. 2021. "Rhythm in the Romance Languages." In *Oxford Research Encyclopedia of Linguistics*. <https://doi.org/10.1093/acrefore/9780199384655.013.431>.
- Brown, Helen. 2013. "Katy Perry, Prism, review." *The Telegraph*, October 17, 2013. <https://www.telegraph.co.uk/culture/music/rockandpopreviews/10386191/Katy-Perry-Prism-review.html>.
- Burzio, Luigi. 1994. *Principles of English Stress*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519741>.
- _____. 2007. "Phonology and Phonetics of English Stress and Vowel Reduction." *Language Sciences* 29 (2-3): 154–76. <https://doi.org/10.1016/j.langsci.2006.12.019>.
- Caplin, William E. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford University Press.
- _____. 2004. "The Classical Cadence: Conceptions and Misconceptions." *Journal of the American Musicological Society* 57 (1), 51–118.
- Carter-Ényi, Aaron, and Gilad Rabinovitch. 2021. "Onset and Contiguity: Melodic Feature Reduction and Pattern Discovery." *Music Theory Online* 27 (4). <https://doi.org/10.30535/mto.27.4.2>.
- Cason, Nia, Corine Astésano, and Daniele Schön. 2015. "Bridging Music and Speech Rhythm: Rhythmic Priming and Audio-Motor Training Affect Speech Perception." *Acta Psychologica* 155: 43–50. <https://doi.org/10.1016/j.actpsy.2014.12.002>.

- Chiu, Matt, and Andrew Blake. 2021. "All the Small Things: Microtiming Deviations in Contemporary Punk." Paper presented at the International Conference on Music Perception and Cognition, virtual conference.
- Chiu, Matt, and Tyler Howie. 2022. "Analytical Frameworks for Post(-Millennial) Punk: We Still [Speak/Sing/Yell] These Songs Well." *SMT-Pod* 1 (12). Audio, 42:22. <https://player.captivate.fm/episode/a552f061-e176-4580-8500-167ab50d6916>.
- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. Harper & Row.
- Claudia1. 2013. "Katy Perry's 'Unconditionally.'" Dis Boards, November 24, 2013. <https://www.disboards.com/threads/katy-perrys-unconditionally.3202871>.
- Coutts, Keppie. 2021. "Can You Really Teach Someone to Write Better Songs?" *HowtoWriteSongs* (blog), June 6, 2021. <https://howtowritesongs.org/2021/06/06/can-you-really-teach-someone-to-write-better-songs/>.
- Covach, John. 2018. "Analyzing Texture in Rock Music: Stratification, Coordination, Position, and Perspective." In *Pop weiter denken: Neue Anstöße aus Jazz Studies, Philosophie, Musiktheorie und Geschichte*, ed. André Doehring and Ralf von Appen, 53–72. Transcript Verlag. <https://doi.org/10.1515/9783839446645-005>.
- Cutler, Anne. 2008. "Lexical Stress." In *The Handbook of Speech Perception*, ed. David B. Pisoni and Robert E. Remez, 264–89. Wiley. <https://doi.org/10.1002/9780470757024.ch11>.
- Daniele, Joseph R., and Aniruddh D. Patel. 2013. "An Empirical Study of Historical Patterns in Musical Rhythm: Analysis of German & Italian Classical Music Using the nPVI Equation." *Music Perception* 31 (1): 10–18. <https://doi.org/10.1525/mp.2013.31.1.10>.
- Degrave, Pauline. 2019. "Music and the Perception of Foreign Language Lexical Stress." PhD diss., Université Catholique de Louvain.
- Deutsch, Diana. 2019. "The Speech-to-Song Illusion: Crossing the Borderline between Speech and Song." In *Musical Illusions and Phantom Words*, 151–69. Oxford University Press. <https://doi.org/10.1093/oso/9780190206833.003.0011>.
- Duinker, Ben. 2021. "Segmentation, Phrasing, and Meter in Hip-Hop Music." *Music Theory Spectrum* 43 (2), 221–245.
- . 2022. "When Hip-Hop Accents Collide (They Create Syncopation)." *SMT-V* 8 (6).
- EW Staff. "Every Katy Perry Song, Ranked." *Entertainment Weekly*, October 25, 2017. <https://ew.com/music/2017/10/25/katy-perry-song-ranking/>.
- Feeney, Nolan. 2013. "Katy Perry's 'Geisha-Style' Performance Needs to Be Called Out." *The Atlantic*, November 25, 2013. <https://www.theatlantic.com/entertainment/archive/2013/11/katy-perrys-geisha-style-performance-needs-to-be-called-out/281805/>.
- Frost, Dan. 2011. "Stress and Cues to Relative Prominence in English and French: A Perceptual Study." *Journal of the International Phonetic Association* 41 (1): 67–84. <https://doi.org/10.1017/S0025100310000253>.
- Fry, D. B. 1955. "Duration and Intensity as Physical Correlates of Linguistic Stress." *The Journal of the Acoustical Society of America* 27 (4): 765–768. <https://doi.org/10.1121/1.1908022>
- Gamboa, Glenn. 2013. "'Prism' Review: Katy Perry is All Over the Place." *Newsday*, October 18, 2013. <https://www.newsday.com/entertainment/music/prism-review-katy-perry-is-all-over-the-place-t97042>.
- Gordon, Reyna L., Cyrille L. Magne, and Edward W. Large. 2011. "EEG Correlates of Song Prosody: A New Look at the Relationship between Linguistic and Musical Rhythm." *Frontiers in Psychology*

2. <https://doi.org/10.3389/fpsyg.2011.00352>.
- Griffiths, Dai. 2003. "From Lyric to Anti-Lyric: Analyzing the Words in Pop Song." In *Analyzing Popular Music*, ed. Allan F. Moore, 39–59. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511482014.003>.
- . 2012. "Internal Rhyme in 'The Boy with a Moon and Star on His Head', Cat Stevens, 1972." *Popular Music* 31 (3): 383–400. <https://doi.org/10.1017/S026114301200030X>.
- Halle, John, and Fred Lerdahl. 1993. "A Generative Textsetting Model." *Current Musicology* 55: 3–23.
- Hausen, Maija, Ritva Torppa, Viljami R. Salmela, Martti Vainio, and Teppo Särkämö. 2013. "Music and Speech Prosody: A Common Rhythm." *Frontiers in Psychology* 4.
<https://doi.org/10.3389/fpsyg.2013.00566>.
- Heffner, Christopher C., and L. Robert Slevc. 2015. "Prosodic Structure as a Parallel to Musical Structure." *Frontiers in Psychology* 6. <https://doi.org/10.3389/fpsyg.2015.01962>.
- Heidemann, Kate. 2016. "A System for Describing Vocal Timbre in Popular Song." *Music Theory Online* 22 (1).
- Ho, Wing See Vincie. 2006. "The Tone-Melody Interface of Popular Songs Written in Tone Languages." Paper presented at the International Conference of Music Perception and Cognition, Bologna, Italy.
- Hoeschele, Marisa, and W. Tecumseh Fitch. 2016. "Phonological Perception by Birds: Budgerigars Can Perceive Lexical Stress." *Animal Cognition* 19 (3): 643–54. <https://doi.org/10.1007/s10071-016-0968-3>.
- Huron, David, and Matthew Royal. 1996. "What is Melodic Accent? Converging Evidence from Musical Practice." *Music Perception* 13 (4): 489–516. <https://doi.org/10.2307/40285700>.
- Isaacson, Eric. 2023. *Visualizing Music*. Indiana University Press.
- Kirby, James, and D. Robert Ladd. 2016. "Tone-Melody Correspondence in Vietnamese Popular Song." Paper presented at Tonal Aspects of Languages, Buffalo, NY.
<https://doi.org/10.21437/TAL.2016-10>.
- Kolinsky, Régine, Hélène Cuvelier, Vincent Goetry, Isabelle Peretz, and José Morais. 2009. "Music Training Facilitates Lexical Stress Processing." *Music Perception* 26 (3): 235–46.
<https://doi.org/10.1525/mp.2009.26.3.235>.
- Komaniecki, Robert. 2019. "Vocal Pitch in Rap Music." Presentation at Society for Music Theory Annual Meeting.
- . 2020. "Vocal Pitch in Rap Flow." *Intégral* 34: 25–46.
- Krebs, Harald. 1999. *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195116236.001.0001>.
- Küster, Martin. 2022. "Should the End of a Phrase Be Emphasized? An Essay in Musical Prosody." In *Topics in Musical Interpretation*, ed. Sezi Seskir and David Hyun-su Kim. Routledge: ebook.
<https://doi.org/10.4324/9781003322269>.
- Ladd, D. Robert, and James Kirby. 2020. "Tone–Melody Matching in Tone-Language Singing." In *The Oxford Handbook of Language Prosody*, ed. Carlos Gussenhoven and Aoju Chen, 676–88. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198832232.013.47>.
- Lee, Christopher S., Lucinda Brown, and Daniel Müllensiefen. 2017. "The Musical Impact of Multicultural London English (MLE) on Speech Rhythm." *Music Perception* 34 (4): 452–81.
<https://doi.org/10.1525/mp.2017.34.4.452>.

- Lerdahl, Fred. 2001. "The Sounds of Poetry Viewed as Music." *Annals of the New York Academy of Sciences* 930 (1): 337–54. <https://doi.org/10.1111/j.1749-6632.2001.tb05743.x>.
- Lerdahl, Fred, and Ray S. Jackendoff. 1983. *A Generative Theory of Tonal Music*. The MIT Press.
- Li, Edwin K.C. 2021. "Cantopop and Speech-Melody Complex." *Music Theory Online* 27 (1).
- Liberman, Mark Yoffe. 1975. "The Intonational System of English." PhD diss., Massachusetts Institute of Technology.
- . 2021. "Operatic Rhetoric." *Language Log* (academic blog), May 21, 2021. <https://languagelog.ldc.upenn.edu/nll/?p=51058>.
- Liberman, Mark, and Alan Prince. 1977. "On Stress and Linguistic Rhythm." *Linguistic Inquiry* 8 (2): 249–336.
- List, George. 1963. "The Boundaries of Speech and Song." *Ethnomusicology* 7 (1): 1–16. <https://doi.org/10.2307/924141>.
- Lynch, Joe. 2016. "Jencarlos Canela Sings Katy Perry's 'Unconditionally' as Jesus on 'The Passion' Live." *Billboard Music News*, March 20, 2016. <https://www.billboard.com/music/music-news/jencarlos-canela-passion-katy-perry-unconditionally-jesus-7263947/>.
- Malawey, Victoria. 2020. *A Blaze of Light in Every Word: Analyzing the Popular Singing Voice*. Oxford University Press.
- McClellan, Laura. 2013. "Katy Perry's New Album in 16 Sentences or More." *Laura McClellan* (personal blog), October 24, 2013. <https://www.lauramcclellan.com/blog/2013/10/katy-perrys-new-album-in-16-sentences-or-more>.
- McCulloch, Gretchen. 2020. *Because Internet: Understanding the New Rules of Language*. Penguin.
- McMurray, Peter. 2021. "Qur'an Alphabetics and the Timbre of Recitation." In *The Oxford Handbook of Timbre*, ed. Emily I. Dolan and Alexander Rehding, 92–120. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190637224.013.1>.
- McPherson, Laura, and Kevin M. Ryan. 2018. "Tone-Tune Association in Tommo So (Dogon) Folk Songs." *Language* 94 (1): 119–56. <https://doi.org/10.1353/lan.2018.0003>.
- Monelle, Raymond. 1989. "Music Notation and the Poetic Foot." *Comparative Literature* 41 (3): 252–69. <https://doi.org/10.2307/1771109>.
- Murphy, Nancy. 2023. *Times A-Changin': Flexible Meter as Self-Expression in Singer-Songwriter Music*. Oxford University Press.
- Neal, Jocelyn R. 2018. "The Twang Factor in Country Music." In *The Relentless Pursuit of Tone: Timbre in Popular Music*, Robert Fink, Melinda Latour, and Zachary Wallmark, 43–64. Oxford University Press. <https://doi.org/10.1093/oso/9780199985227.003.0003>.
- Newman, Melinda. 2013. "Katy Perry's 'Prism' Reflects Her Heart and Soul Set to a Beat: Review." *Uproxx*, October 21, 2013. <https://uproxx.com/hitfix/katy-perrys-prism-reflects-her-heart-and-soul-set-to-a-beat-review/>.
- Nobile, Drew. 2020. *Form as Harmony in Rock Music*. Oxford University Press.
- Ohio State University Department of Linguistics. 2023. *Language Files: Materials for an Introduction to Language and Linguistics*. Ohio State University Press.
- Ohriner, Mitchell. 2019a. "Lyric, Rhythm, and Non-alignment in the Second Verse of Kendrick Lamar's 'Momma.'" *Music Theory Online* 25 (1). <https://doi.org/10.30535/mto.25.1.10>.
- . 2019b. *Flow: The Rhythmic Voice in Rap Music*. Oxford University Press.

- Palmer, Caroline, and Michael H. Kelly. 1992. "Linguistic Prosody and Musical Meter in Song." *Journal of Memory and Language* 31 (4): 525–42. [https://doi.org/10.1016/0749-596X\(92\)90027-U](https://doi.org/10.1016/0749-596X(92)90027-U).
- Palmer, Caroline, Melissa K. Jungers, and Peter W. Jusczyk. 2001. "Episodic Memory for Musical Prosody." *Journal of Memory and Language* 45 (4): 526–45. <https://doi.org/10.1006/jmla.2000.2780>.
- Patel, Aniruddh D. 2010. *Music, Language, and the Brain*. Oxford University Press.
- Pattison, Pat. 2014. "Wittgenstein, Ordinary Language, and Songwriting." *Fusion*, December 10, 2014. <https://www.fusionmagazine.org/wittgenstein-ordinary-language-and-songwriting-2/>.
- Pau, Andrew. 2015. "'Sous le rythme de la chanson': Rhythm, Text, and Diegetic Performance in Nineteenth-Century French Opera." *Music Theory Online* 21 (3). <https://doi.org/10.30535/mto.21.3.9>.
- Perricone, Jack. 2018. *Great Songwriting Techniques*. Oxford University Press.
- Pike, Kenneth L. 1945. *The Intonation of American English*. University of Michigan Press.
- Povel, Dirk-Jan, and Peter Essens. 1985. "Perception of Temporal Patterns." *Music Perception* 2 (4): 411–40. <https://doi.org/10.2307/40285311>.
- Proto, Teresa. 2015. "Prosody, Melody and Rhythm in Vocal Music: The Problem of Textsetting in a Linguistic Perspective." *Linguistics in the Netherlands* 32 (1): 116–29. <https://doi.org/10.1075/avt.32.09pro>.
- Provenzano, Catherine. 2018. "Auto-Tune, Labor, and the Pop-Music Voice." In *The Relentless Pursuit of Tone: Timbre in Popular Music*, ed. Robert Fink, Melinda Latour, and Zachary Wallmark, 159–82. Oxford University Press. <https://doi.org/10.1093/oso/9780199985227.003.0008>.
- Reed, Alissandra, Braden Maxwell, and David Temperley. 2019. "P2-44 Cognitive Coupling of Stress and Meter." Poster presented at Society for Music Perception and Cognition meeting, August 10, 2019, New York, NY.
- Rodgers, Stephen. 2015. "The Fourth Dimension of a Song." *Music Theory Spectrum* 37 (1): 144–53. <https://doi.org/10.1093/mts/mtv002>.
- Rodríguez-Vázquez, Rosalía. 2010. "Text-Setting Constraints: A Comparative Perspective." *Australian Journal of Linguistics* 30 (1): 19–34. <https://doi.org/10.1080/07268600903134004>.
- Rothstein, William Nathan. 1989. *Phrase Rhythm in Tonal Music*. Schirmer Books.
- . 2008. "National Metrical Types in Music of the Eighteenth and Early Nineteenth Centuries." In *Communication in Eighteenth-Century Music*, ed. Danuta Mirka and Kofi Agawu, 112–59. Cambridge University Press.
- Salley, Keith. 2011. "On the Interaction of Alliteration with Rhythm and Meter in Popular Music." *Popular Music* 30 (3): 409–32. <https://doi.org/10.1017/S0261143011000225>.
- Santa_Bond69. 2020. "Katy Perry's Pronunciation of "Unconditionally" as UN-con-dish-SHUN-a-llly" Instead of "un-con-DISH-SHUN-ally" Irrationally Angered and Ruined the Song for Me..." Reddit (post thread), December 22, 2020. https://www.reddit.com/r/popheads/comments/ki8pz2/katy_perrys_pronunciation_of_unconditionally_as/.
- Schellenberg, Murray. 2012. "Does Language Determine Music in Tone Languages?" *Ethnomusicology* 56 (2): 266–78. <https://doi.org/10.5406/ethnomusicology.56.2.0266>.
- Schreuder, Maartje. 2006. "Prosodic Processes in Language and Music." PhD diss., Rijksuniversiteit Groningen, School of Behavioral and Cognitive Neurosciences.
- Sloan, Nate and Charlie Harding. 2019. "Shawn Mendes and the Rhythm of Rhyme." *Switched on Pop* (podcast) 125. <https://switchedonpop.com/episodes/shawn-mendes-and-the-rhythm-of-rhyme>.

- Tait, Casey, Marija Tabain, and Ingrid Sykes. 2014. "Stress-Meter Alignment in American Hip Hop." Paper presented at 15th Australasian International Conference on Speech Science and Technology, Christchurch.
- Temperley, David. 1999. "Syncopation in Rock: A Perceptual Perspective." *Popular Music* 18 (1): 19–40. <https://doi.org/10.1017/S026114300008710>.
- . 2007. "The Melodic-Harmonic 'Divorce' in Rock." *Popular Music* 26 (2): 323–42. <https://doi.org/10.1017/S0261143007001249>.
- . 2019. "Second-Position Syncopation in European and American Vocal Music." *Empirical Musicology Review* 14 (1-2): 66–80. <https://doi.org/10.18061/emr.v14i1-2.6986>.
- . 2021. "The Origins of Syncopation in American Popular Music." *Popular Music* 40 (1): 18–41. <https://doi.org/10.1017/S0261143021000283>.
- . 2022. "Music and Language." *Annual Review of Linguistics* 8: 153–70. <https://doi.org/10.1146/annurev-linguistics-031220-121126>.
- Temperley, Nicholas, and David Temperley. 2011. "Music-Language Correlations and the 'Scotch Snap.'" *Music Perception* 29 (1): 51–63. <https://doi.org/10.1525/mp.2011.29.1.51>.
- . 2013. "Stress-Meter Alignment in French Vocal Music." *The Journal of the Acoustical Society of America* 134 (1): 520–27. <https://doi.org/10.1121/1.4807566>.
- Thomas, Margaret E. 2016. "Text and Temporality: Toward an Understanding of Rhythmic Irregularities in the Music of Tom Waits." *Music Theory Online* 22 (4). <https://doi.org/10.30535/mto.22.4.5>.
- Trehub, Sandra E., Judith Becker, and Iain Morley. 2015. "Cross-Cultural Perspectives on Music and Musicality." *Philosophical transactions of the Royal Society of London: Series B* 370 (1664). <https://doi.org/10.1098/rstb.2014.0096>.
- TVTropes. Last modified April 11, 2024. "AcCENT upon the Wrong SylLABLE." Accessed February 27, 2023. <https://tvtropes.org/pmwiki/pmwiki.php/Main/AcCENTUponTheWrongSylLABLE>.
- VanderStel, Joseph. 2021. "The Evolution of Syncopation in 20th-Century Popular Music." PhD diss., University of Rochester.
- Varwig, Bettina. 2021. "Early Modern Voices." In *The Oxford Handbook of Timbre*, ed. Emily I. Dolan, and Alexander Rehding, 249–267. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190637224.013.32>.
- Vukovics, Katherine, and Daniel Shanahan. 2020. "Rhythmic Variability, Language, and Style: A Replication and Extension of nPVI Findings with the RISM Dataset." *Journal of New Music Research* 49 (3): 285–97. <https://doi.org/10.1080/09298215.2020.1751209>.
- Wennerstrom, Ann. 2001. *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford University Press. <https://doi.org/10.1093/oso/9780195143218.001.0001>.

Footnotes

* To reference this paper, please use she/her or they/them pronouns and cite the author by first name (e.g. Eron 2024).

[Return to text](#)

1. Thank you to my colleague and friend Suraj Saifullah for providing me with this example.
[Return to text](#)

2. For several anecdotal examples, see the Disboards forum by Claudia1, the r/popheads Reddit thread by Santa_Bond69, the song's listing in TVTropes, reference to "syllabic phrasing" in

Entertainment Weekly's ranking of all of Katy Perry's songs, Coutts (2021 blog post), Gamboa (2013 Newsday review), McClellan (2013 blog post), Newman (2013 Uproxx review), and Pattison (2014). A search for "syllable emphasis unconditionally katy perry" on Twitter also yields many representative judgments.

[Return to text](#)

3. For a helpful review of many of the concepts below, see *Music, Language, and the Brain* (Patel 2010) and the Lexical Stress chapter in *The Handbook of Speech Perception* (Cutler 2008).

[Return to text](#)

4. Some scholarship, such as Heffner and Slevc (2015), Schreuder (2006), and Temperley (2022), has explicitly highlighted these commonalities between music and language.

[Return to text](#)

5. These features of English stress, as well as the stress- vs. syllable-timed language dichotomy, are widespread in mid-20th-century linguistics; once again, see Patel (2010). Representative primary sources include Pike (1945), generally considered to be the original source of the stress/syllable dichotomy; Burzio's influential *Principles of English Stress* (1994); Chomsky & Halle (1968), who not only offer guidelines on word pronunciation but also methods for quantifying stress; and Fry (1955), who found that duration is a better indicator of stress than intensity on spectrograms.

[Return to text](#)

6. Some other languages have features that English does not, such as tones, which are more influential to texted music in those languages. This is well documented; for representative examples, see Kirby and Ladd (2016), McPherson and Ryan (2018), and Proto (2015).

[Return to text](#)

7. List (1963) provides examples of many of these cultural distinctions.

[Return to text](#)

8. For a sampling of recent research on this topic, see Adams (2008; 2009), Duinker (2022), Ohriner (2019a, 2019b), Patel (2010), Salley (2011), Temperley (1999), and VanderStel (2021).

[Return to text](#)

9. As a demonstration of this assumption, see Lee, Brown, and Müllensiefen (2017); Palmer and Kelly (1992); Patel (2010); Proto (2015); Rodríguez-Vásquez (2010); Tait, Tabain, and Sykes (2014); and Temperley and Temperley (2011; 2013). The latter remarks, in particular, that the "lack of interaction between the two sources of rhythmic structure" is interesting (539).

[Return to text](#)

10. For examples of these visualizations and hierarchical structures, see Aroui and Arleo (2009), Halle and Lerdahl (1993), Lerdahl (2001), Liberman (1975), Liberman and Prince (1977), Monelle (1989), and Patel (2010).

[Return to text](#)

11. Five representative examples of text as an analytical venue may be found in BaileyShea (2021), Griffiths (2003), Pau (2015), Rodgers (2015), and Thomas (2016).

[Return to text](#)

12. Pedagogical resources past and present consistently emphasize the need for alignment between prosody and melody. Küster (2022) returns to rhetoric-centered music theories of the 18th century to suggest that prosody should factor into performance of phrase, while 21st-century songwriting instruction in textbooks and blog posts alike warn against the type of mismatches in Example 1 above (e.g. Perricone 2018 and Pattison 2014).

[Return to text](#)

13. As to whether these conflicts affect quality, I believe value judgments lie beyond the scope of responsible music scholarship.

[Return to text](#)

14. I use “dissonance” here in the same sense of misalignment as Krebs (1999); *prosodic consonance* would be the alignment of these features. This article focuses primarily on stress, which is the main connotation of “prosody” for English; which elements are phonemic or suprasegmental depends on the language (see fn. 39 below) and would redefine prosodic dissonance accordingly.

[Return to text](#)

15. Of course, other languages also have lexical categories and semantic content, and other spoken languages also have vowels.

[Return to text](#)

16. “Musical stress” refers in almost all cases to melodic stress, with some exceptions (see Example 15).

[Return to text](#)

17. Stressing a spoken phrase to change the meaning is sometimes called “contrastive stress.”

[Return to text](#)

18. My own dialect will inform all of my subsequent analyses: I am a white American who grew up primarily in the urban southeast and went to college in southern California.

[Return to text](#)

19. This is supported by Carter-Ényì and Rabinovitch (2021), Huron and Royal (1996), and Lerdahl and Jackendoff (1983).

[Return to text](#)

20. Sloan and Harding (2019) use this example in their discussion of the declamation of “somebody.”

[Return to text](#)

21. This describes the well-known concept of “durational accent,” as discussed in Lerdahl and Jackendoff (1983). For a representative citation of the perceptual effect, see Povel and Essens (1985).

[Return to text](#)

22. As noted in Patel 2010 (176–77), general isochrony in speech is not empirically supported. However, we can still say with confidence that Olivia Rodrigo’s isochronous melody clashes more actively with an idiomatic, “maximally even” distribution of spoken syllables compared to Dua Lipa’s.

[Return to text](#)

23. This is considered general knowledge in American linguistics; see Burzio (2007) and Patel (2010).

[Return to text](#)

24. A schwa is the relaxed “uh” sound, as in the first syllables of “ago” or “pathetic” in many dialects of English.

[Return to text](#)

25. Cutler (2008) provides a handy review of some of these effects.

[Return to text](#)

26. Interestingly, this type of prosodic dissonance is more widely accepted as an expressive choice rather than a mistake by fans of the genre, to the point that rap and hip-hop are discussed on the Wikipedia page for “internal rhymes” and fans post on forums asking for suggestions of rappers that use internal and slant rhymes. For a representative analysis of internal rhyme, see Griffiths (2012). Of course, my intention is not to add a negative value judgment to expressive rhyme creation and emphasis, but to highlight that prosodic dissonance has already entered mainstream aesthetic values in some contexts.

[Return to text](#)

27. See for example Adams (2009).

[Return to text](#)

28. Janelle Monáe’s musical rendition of “compression” also reinforces the sexual imagery prevalent throughout the song.

[Return to text](#)

29. This rhyme-emphasis prosodic dissonance is, as mentioned below, a common feature of rap music, to the point that it does not seem to convey “mispronunciation” in the same way.

[Return to text](#)

30. VanderStel and Temperley both use the term “lexical syncopation.”

[Return to text](#)

31. *Tresillo* here means a 3+3+2 grouping, popularized in Afro-Cuban music and often used in mainstream American pop genres (particularly in dance styles and American-radio Latin music).

[Return to text](#)

32. Additional support for the prosodically dissonant reading comes from the resulting emphasis on “art” and “pop,” given that the name of this album is *Artpop*.

[Return to text](#)

33. Establishing an empirical basis for these two competing modes of listening is beyond the scope of this project. I would expect that listeners would default to a syncopated hearing in an ambiguous case, and that experiencing syncopation rather than prosodic dissonance would be more easily available in common metric positions for syncopation (e.g., second- and fourth-position syncopation, see Temperley 2019; 2021).

[Return to text](#)

34. Thanks to Kyle Adams for suggesting an explicit connection between the minimal pitch emphasis in rap and the rarity of prosodic dissonance.

[Return to text](#)

35. A full literature review of the definition of “phrase” is beyond the scope of this article, but interested readers can trace the definition of phrase through a variety of sources and musical styles including Attas (2011), Caplin (1998; 2004), Duinker (2021), Murphy (2023), Nobile (2020), Ohriner (2019b), and Rothstein (1989). Phrase perception is also repeatedly a subject of music cognition studies.

[Return to text](#)

36. Pitch, or, more specifically, intonation, plays a role in delineating phrase boundaries in spoken English as well; however, since this differs widely across dialects, I restrict my purview here to the timing component of phrase marking.

[Return to text](#)

37. This phrase break also emphasizes the syllable “-peat,” which could be interpreted as a nod to the singer’s breakup with ex-fiancé Pete Davidson. (Thanks to my anonymous peer reviewer for pointing this out.)

[Return to text](#)

38. In addition to the rap scholarship mentioned in the introduction, particularly Duinker (2022), any analysis of accent and rap bears mentioning Mitch Ohriner’s book (2019b). This particular verse could be well described by a 222332 vocal groove (thanks to Drew Nobile for suggesting this connection).

[Return to text](#)

39. The first line actually differs slightly in having a different pickup, but the lines are identical in all other respects.

[Return to text](#)

40. Once again, note that there *is* some syncopation in the initial couplet, but only at the very end of each line—contrasting with the *tresillo* beginning on a weak beat in line 4.

[Return to text](#)

41. Rather than transcribe this verse, I have chosen to represent it through line breaks in the text.

[Return to text](#)

42. Interestingly, this suggests one way enjambment dissonance differs from accent-based prosodic dissonance, since its representations sometimes find their way into public-facing representations.

[Return to text](#)

43. Contrast this setting of “don’t know” with Taylor Swift’s in Example 4 above—note how the same words might be stressed differently in a different speech context.

[Return to text](#)

44. Rina’s 2022 performance on *The Tonight Show* provides a good example:

<https://www.youtube.com/watch?v=zON14vKm464>.

[Return to text](#)

45. This consonant interpretation requires some emphasis on “what,” which doesn’t sound unidiomatic but does evoke some contrastive stress: “don’t know *what* I did. . .”

[Return to text](#)

46. It would still be difficult to hear “saw *a*” as dissonant, however, since the weakest syllable still lands on beat 2 in contrast to the stronger syllables landing on beats 1 and 3.

[Return to text](#)

47. I credit James Sullivan (personal correspondence) with the idea that formal sections might correlate with instances of prosodic dissonance.

[Return to text](#)

48. The correspondence of melodic pitch and phonemic tone has been previously investigated by a variety of scholars; see for example Ho (2006), Kirby and Ladd (2016), Ladd and Kirby (2020), Li (2021), McPherson and Ryan (2018), and Schellenberg (2012).

[Return to text](#)

49. For some representative examples, see Heidemann (2016), Malawey (2020), McMurray (2021), Neal (2018), Provenzano (2018), and Varwig (2021).

[Return to text](#)

50. For an earlier version of this paper, I had the chance to correspond with Sara Albert on her undergraduate capstone project, which investigates the effect of song familiarity on pronunciation of the corresponding lyrics (Albert 2021).

[Return to text](#)

51. Sloan and Harding (2019) provide examples of declamation of the word “somebody” in particular (with gratitude to Drew Nobile for the recommendation).

[Return to text](#)

52. McCulloch (2020) provides an accessible introduction to internet linguistics, including some of these visual effects.

[Return to text](#)

53. Isaacson (2023) provides a comprehensive survey of visualization techniques in music as well as an in-depth exploration of what these techniques illustrate. Combining music visualization with linguistic/text visualization techniques would be a fruitful and interesting future research project.

[Return to text](#)

54. This is generally accepted in current linguistics; for a representative example see *Language Files*, a 2023 textbook published by the linguistics department at The Ohio State University, which was

my personal introduction to descriptivism.

[Return to text](#)

55. For example, if English speakers were to gravitate increasingly toward saying “____ and me went to school” rather than the grammatically prescribed “____ and I went to school,” whether this is “wrong” has little to do with investigating how this change came about or what it says about English speakers’ hearing of case.

[Return to text](#)

56. This reading is supported further by the prosodic dissonance on “dirty laundry” in the first verse, which seems to pair the listener’s flaws directly with shameful secrets—and, I would argue, undermined by Katy Perry’s offensive, aggressively “Japanese” performance of the song at the 2013 AMAs, which feels to me incongruent with the message of unconditional love.

[Return to text](#)

Copyright Statement

Copyright © 2024 by the Society for Music Theory. All rights reserved.

[1] Copyrights for individual items published in *Music Theory Online* (*MTO*) are held by their authors. Items appearing in *MTO* may be saved and stored in electronic or paper form, and may be shared among individuals for purposes of scholarly research or discussion, but may *not* be republished in any form, electronic or print, without prior, written permission from the author(s), and advance notification of the editors of *MTO*.

[2] Any redistributed form of items published in *MTO* must include the following information in a form appropriate to the medium in which the items are to appear:

This item appeared in *Music Theory Online* in [VOLUME #, ISSUE #] on [DAY/MONTH/YEAR]. It was authored by [FULL NAME, EMAIL ADDRESS], with whose written permission it is reprinted here.

[3] Libraries may archive issues of *MTO* in electronic or paper form for public access so long as each issue is stored in its entirety, and no access fee is charged. Exceptions to these requirements must be approved in writing by the editors of *MTO*, who will act in accordance with the decisions of the Society for Music Theory.

This document and all portions thereof are protected by U.S. and international copyright laws. Material contained herein may be copied and/or distributed for research purposes only.

Prepared by Lauren Irschick, Editorial Assistant

